DOCUMENT RESUME

ED 194 594 TM 800 724

AUTHOR vanderPloeg, Arie J.
TITLE Formal and Informal Evaluation: To Serve Two Masters Well.
PUB DATE Apr 80
NOTE 22p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Compensatory Education: *Educational Assessment: Elementary Secondary Education: *Evaluation Methods: *Program Evaluation
IDENTIFIERS Chicago Public Schools IL: *Elementary Secondary Education Act Title I

ABSTRACT

In evaluation of the Elementary Secondary Education Act Title I compensatory education programs, both the federal reporting mandate (using formal evaluation) and local needs (using informal evaluation) must be served. In Chicago Public Schools, computerized processing of standardized test scores allows the reporting requirement to be met with a minimum of staff and time. The bulk of staff and time are used to gather information about program implementation and conduct at the local schools. This information is returned to school principals and program administrators several times yearly. The local evaluation is beginning to focus on two sets of persons as its subject: (1) the student-teacher relationship which sets the limits to pupils' learning: and (2) the school principal and the various administrators who have the power to promote or vitiate programs. The reality in which these persons act fluctuates. Nevertheless, that varying reality is the context in which decisions about Title I programs are made. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194594

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FORMAL AND INFORMAL EVALUATION: TO SERVE TWO MASTERS WELL

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Vande Ploeg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Arie J. vanderPloeg
Department of Research, Evaluation and Long Range Planning
Chicago Public Schools
2021 North Burling Street
Chicago, IL 60614

Paper presented at the annual meeting of the
American Educational Research Association,
Boston, 9 April 1980

TM 800724

FORMAL AND INFORMAL EVALUATION
TO SERVE TWO MASTERS WELL

The form and content of educational evaluation have recently been much criticized, from within and without the profession. Policy makers and administrators contend that educational evaluation is of poor quality, lacks credibility, fails to address relevant issues, and arrives at the wrong moment (Bissell 1979; Florio, Behrmann and Goltz 1979; Wisler and Anderson 1979). Some practitioners claim that the dominant mode of evaluation fails to provide appropriate descriptions, that emphasis on the "straight facts" distorts the meaning of events (Eisner 1979a; Schwab 1970; Stake 1976.) Others argue the technical details of alternative representations of the "facts" (Burstein and Miller 1978; Linn and Slinde 1977.)

In part these criticisms reflect the individual perspectives, the different audiences, and the varying information needs of the commentators. Nevertheless, a fundamental question underlies the concerns raised. This has to do with the adequacy of the dominant tradition of educational research, that based on the desire to use the methods of science to explicate social phenomena. Great technical strides have been made within this tradition and rigorous canons of ethics and procedure control and protect the use of these methodologies. Still, most educational evaluators--at least those I know--share Eisner's (1979b, p.11) uneasiness:

> ... such approaches, and the methods of inquiry that are regarded as legitimate within their borders, somehow fail to tell the whole story. As a result of the partial view that such methods provide, a biased, even distorted picture of the reality that we are attempting to understand and improve can occur.

The RMC models for the evaluation of ESEA Title I compensatory education programs (Tallmadge and Wood 1976) are the specific examples responsible for the unease about the form, content, and meaning of educational evaluation which I wish to discuss. These models aim to provide federal policymakers with useful information regarding the effectiveness of a myriad of special educational activities implemented by local schools. Those who devised these models appear to have understood the complexity of the task assigned to them; nevertheless, they remain controversial (Gamel, Tallmadge, Wood and Binkley 1975; Kearns 1978; Linn 1979; Wiley 1979.)

Title I programs are by definition local. Now matter how technically adequate, the upward aggregation of data must result in a loss of information.[1] The critical question is: are the data which survive the aggregation process meaningful? Since these are compensatory education programs, an indicator of improvement based on standardized achievement test scores--the much discussed Normal Curve Equivalent gains--was developed. Yet, at the local level, test scores are not always the most meaningful indicators of project impact. Knowledge of any group of pupils' gain scores is an inadequate indicator of program effect, as the local evaluator and administrator are all too well aware.

To quote Eisner again:

> ... There is, of course, nothing wrong with knowing how well or how poorly a student performs. Yet schools, insofar as they are educational institutions, should not be content with performance. Education as a process is concerned with the cultivation of intellectual power, and the ability to determine what a student knows is not necessarily useful or sufficient for making that process more effective.

There is a second cause for concern. The emphasis of the Title I evaluation has been on the implementation of the national reporting system. This has caused local evaluations to focus their attention on the technical and bureaucratic problems of obtaining and reporting data to state education agencies. This emphasis may have resulted in too little thought and too little manpower being available for the local aspects of the Title I evaluation. Yet, any evaluation of programs that are locally designed and implemented must emphasize local conditions and circumstances to be meaningful. A brief history of Title I evaluation in Chicago may illustrate this.

The Title I project in Chicago is now more than a dozen years old. It annually serves almost 70,000 pupils in about 300 schools in a district composed of approximately 600 elementary schools enrolling roughly half a million pupils. Over 40 distinct educational activities are provided with Title I monies. Schools select from among these activities as their pupils' needs and their funds permit. The initial emphasis of the local evaluation was on the effectiveness of the individual activities, paralleling the national emphasis

on project evaluations. Three phases are evident in the history of Title I evaluation in Chicago.

The initial evaluation effort made use of outside contractors to evaluate the individual activities. The function of the Board of Education's Title I evaluators, of which there were usually about eight, was principally one of liaison between the external contractors and local school and central office staff. The external contractors carried out almost all instrument development, data collection (with the exception of achievement testing which was done on a city-wide basis by the Board), and report writing. As a consequence, Board personnel were only indirectly involved with the specific Title I programs and the schools in which they operated.

The evaluation reports produced in this manner were of greatly varying quality. Given numerous external contractors, uniformity of reports was extremely limited and comparisons among programs generally impossible. Publication was frequently delayed: in some cases reports were not available until several years after the close of a school year. Duplication of data collection procedures was common. Rectification of procedural errors was difficult. And not least, the contracted evaluators were often far removed from the day-to-day developments in schools and programs, forcing reports of such generality that their utility became questionable. The required reporting to the state educational agency, meanwhile, took the form of transmitting a computer tape of the test results for Title I pupils obtained as a part of the annual city-wide spring testing. Analyses of these data at the local level

was originally minimal but came with time to be the primary thrust of the evaluation.

In part in response to the obvious inadequacy of these procedures, the Department of Research and Evaluation was formed as a semi-independent branch of the central administration, reporting directly to the general superintendent. As quickly as possible, the practice of using outside contractors was discontinued. Eight evaluators were assigned to a Title I evaluation unit. Each evaluator was responsible for the evaluation of five or six Title I activities. The Department also instituted a series of innovations designed to expedite data collection, improve comparability, provide more detailed and comprehensive evaluation reports, and ensure more prompt evaluations. Data collection forms were reduced in number. Two page interim reports--necessitated by the continued delay of the final reports--were provided to assist schools in the selection of activities. Considerable financial savings were also realized. For instance, in 1976-77 the externally contracted Title I classroom observation study cost roughly $40,000. The following year, using Board of Education staff, the cost of this aspect of the evaluation was reduced to under $15,000. The time for the evaluation effort decreased from one year to less than six months.

The central purpose of the evaluation remained the identification of the most effective activities as determined by standardized test results. This task required the comparison

of activities enrolling pupils of varying age groups and ability levels. A method of transforming achievement test scores, similar to that propounded in the RMC models but developed independently, was chosen. Within age-cycle distributions of pretest and posttest scores were individually standardized and differences in the distributional placement of individuals at pretest and posttest times were compared. These data confirmed that in terms of test performance most activities produced approximately the same degree of pupil growth. Each year only two or three activities could be identified to be more effective than others. There was little consistency from year to year concerning relative effectiveness: a program appearing effective one year often demonstrated only average results the following year.

From this a set of intertwined conclusions was drawn: the efficiency of Title I activities in Chicago depends more upon which schools select activities, which pupils schools choose to participate, how individual schools implement activities, and the amount of staff training the program vendors supply. Each of these conditions could and did change from year to year. Regardless of the precision with which an activity was specified, schools could be counted upon to modify the implementation of an activity along lines most appealing to the local school staff. Clearly, under such conditions the focus on pupils' academic achievement within activities could not be expected to provide a helpful evaluation.

In consequence, the Department in 1978-79 revised the evaluation effort once more. Two principal improvements were sought: near immediate reporting and school level evaluation. Nine Title I evaluators were each assigned to a specific set of schools. Most of their time was spent in these schools observing and assisting local school Title I staff. The evaluators were required to prepare anecdotal summaries of each visit, to obtain certain minimal information on program enrollments and characteristics, and to conduct classroom observations and staff interviews. Given a sufficient level of professionalism and trust, it was expected that the field evaluators would be able to obtain evaluative information even as they provided feedback to the local schools in the attempt to improve the conduct of Title I instruction. At the Department of Research and Evaluation a staff of three or four persons monitored field events, coordinated and interpreted the information received, wrote the final reports, and prepared brief reports on special topics for immediate dissemination to the schools and district staff.

From the point of view of the Department, this approach greatly improved the Title I evaluation. Information on the local schools and the Title I activities they selected has become much more detailed and annual and periodic reports seem much more informative and useful. School and central office administrators appear to concur in this assessment (Hamilton 1980). Nevertheless, problems remain.[2] Not the least of these is tension within the staff concerning the

proper role of evaluation. Some see a role based on evaluation of "hard" data. They interpret the federal mandate and the RMC models as the primary responsibility. The majority contend this duty is a minimal requirement and that a more thorough interpretation and understanding must be reached.

Although I have no wish to add another neologism, let alone one that is dichotomous, to the stock of jargon current in educational evaluation, I would like to make use of a pair of labels as intellectual shorthand. The nature of the uneasiness I have referred to can best be understood by distinguishing between formal and informal evaluation. This dichotomy is in many ways similar to others adumbrated in the recent past: Schermerhorn and Williams' (1979) preordinate and responsive approaches, La Belle, Moll and Weisner's (1979) juxtaposition of input-output and participant researcher strategies, Schwab's (1970) theoretic and practical orientations, Getzels and Guba's (1957) distinction between the nomothetic and the idiographic, summative versus formative evaluation (Scriven 1967), to name a few. I will make no effort to spell out precisely the philosophical differences nor trace the history of these concepts in the sociology of knowledge of evaluation, although elements of the tension between the empirical and phenomenological schools of social analysis are evident. The empiricist attempts to uncover social "truths" and uses them to determine consequences; the phenomenologist is more interested in comprehending "meanings". The former holds "laws" to be the primary object of inquiry;

the latter assigns priority to persons and intentions. I wish to emphasize that I do not intend to denigrate any method of analysis or school of evaluation; what I hope to accomplish is a specification of usefulness and proper place.

Formal evaluation stresses scientific methods and numerical technology. It main tools are descriptive and inferential statistics and the discipline of experimental design and analysis. The intent of formal evaluation is the uncovery of reproducible educational outcomes, particularly student outcomes. It aspires to objectivity and neutrality. Formal evaluation reports, while they may be descriptive, prefer to elicit the laws which govern behavior. In this attempt sophisticated statistical methods are called upon.[3] The complement these methods, equally sophisticated measurement of the traits of interest are sought. Since distributions of students' scores on standardized tests often have most of the properties these methods require, they are frequently the subject of formal evaluation. Conclusions are generally stated with hesitancy and reticence and are much hemmed about with qualification. Formal evaluation is the dominant mode in educational research and evaluation today, dominant both in being the professional ideal to which most evaluators strive and in being the form most evaluation reports take.

Although many evaluators now call for evaluations designed to improve educational processes, the history of

evaluation and the training of most evaluators have emphasized summation, classification, and technique. Despite current rhetoric, the 1971 statement of Bloom, Hastings, and Madaus remains true: "As testing and other forms of evaluation are commonly used in the schools, they contribute little to the improvement of teaching and learning" (p. 7). The evaluator's own desires for advancement urge him to use formal methods: the professional journals appear more likely to publish papers displaying traditional research designs and sophisticated numerical skills. Paradoxically, it is often easier to author a statistically sophisticated report than to write a clear and meaningful evaluation using logic, vocabulary, and the rules of grammar as tools.

The RMC models for the evaluation of Title I programs fall squarely within this tradition of formal evaluation. Given the concerns of administrators and legislators at the federal level, this may be appropriate. Summative results appear to be useful for yes-or-no decisions regarding funding or its continuation.

At the local level emphasis is on improving instruction, usually by modifying programs already in place. Local resources, for a variety of reasons, simply do not permit continual picking and choosing among programs and curricula. Informality, improvement, feedback, formative evaluation--these are the lifeblood of local evaluation. The RMC models do little to improve circulation; they are designed to indicate only whether no or major surgery is required. More generally,

the formal empiricist approach has limited utility locally. Often properly trained staff are unavailable. Outside contracts are expensive and as in Chicago, may not be able to capture the flavor of local events. A more informal, phenomenological approach may be more suitable to local needs. Events have a way of marching onward, oblivious to the needs of the researcher. Formal evaluation has little capacity to adapt to the succession of events. Informal approaches, however, can capitalize on this.

I would like to illustrate these concerns using the progression of evaluation schema just described for Chicago. The person most capable of exerting control over the implementation and modification of instruction is the local principal. Second in influence is probably the Title I coordinator, whose task it is to assist Title I teachers. Much further down the list is the local evaluator. While he often feels he has information of great usefulness to principals and coordinators, he finds himself ignored and ineffectual. Why?

The Title I evaluator operates on a basis much different from that which guides the principal. In many ways the principal is isolated from the circumstances determining the behavior and orientation of central office staff. He understands his school, his staff, and his students. Principals tend to share the conviction that they know their own territory best. Orders, requests, even information from other sources in the educational bureaucracy all are, to varying extents, unwelcome infringements on the business of running a school. Evaluation

is perceived to be legitimated by outside sources. Principals quickly perceive that much of the testing and form-filling is not intended to help them, but to fulfill other obligations. Catering to the evaluator's needs is equivalent to obeying outside directives. This reduces the principal's perception of his own authority. Especially as regards Title I, where the major justification for evaluation is a federal or state requirement, this produces only grudging cooperation. Reports of personal and professional recalcitrance were not infrequently made by the external contractors in the first phase of Chicago's Title I evaluation. With the switch to school-focused evaluation such complaints still arise, though more often based on personality than on professional differences.

These tendencies are exacerbated by emphasis on formal modes of evaluation. Standardized testing of pupils and completion of rigid questionnaires by staff limit the principal's autonomy and authority. He cannot control the responses his pupils or staff make. Even if he wished wholeheartedly to improve pupil performance on tests, he knows that current educational technology is insufficient to cause dramatic change. His pupils' test performance, particularly from year to year, is essentially fixed by the quality of the students it is his lot to lead. He knows he can influence other characteristics of his students and his school, but these are not usually measured or evaluated, certainly not wherever the RMC models set the tone for the evaluation. Despite his best efforts, the evaluation, as he sees it, will make conclusions and recommendations which he can do little to implement.

Chicago's Title I evaluation reports at first emphasized the interpretation of mean scores of individual Title I activities but said little about the performance of pupils at individual schools. Given that a principal was convinced that a particular activity worked well at his school, he was likely to shrug off a report that the activity mean was below average. After all, some schools must be below and others above the average. So what if most schools reported difficulty in implementing Computer-Assisted Instruction? Once the computers were properly connected, the CAI classroom at his school became an active, happy learning place.

To the extent analysis of data is done along formal lines, the principal may not even be able to comprehend the methods the evaluator uses to come to his conclusions. Such technical intimidation cannot cause him to react favorably. Most principals are trained by their experience to mistrust what they cannot understand. Talk of NCEs and multivariate regression is foreign to him. It does not appear to relate to more urgent problems: How can I improve teacher X's instruction? What attitude do I take to the parents of Johnny to get him to school more regularly? In Chicago, staff of the Department of Research and Evaluation might well appreciate the rigor and elegance of some consultant's analysis, but in the schools the analysis was often incomprehensible and therefore threatening and untrustworthy.

Clearly this litany could be extended. Nor is it unfamiliar. The standard panacea, often espoused by the Technical Assistance Centers created by the ESEA legislation, is to inservice, involve, and educate local staff. Given time, this may work--especially if the TACs shift their focus from the local evaluator to the local principal. In the meantime, informal approaches to evaluation may create a better environment and more meaningful evaluations.

By advocating more informal evaluations, I do not mean to suggest that proven techniques be abandoned or that the rigor with which evaluations are conducted be diluted. Rather, greater involvement in evaluation, particularly in evaluation design and interpretation, should be sought from those persons who have a real stake in the success of Title I programs: the local principal, coordinator, and program vendor. The RMC models should not intrigue the local evaluator so much that he retains too little enthusiasm for providing information useful to local principals and others with the capability to improve the content of instruction in Title I.

Informal evaluation requires the evaluator to take a more creative, more critical stance than does formal evaluation. Neither, to be effective, permits sloppy workmanship. The call for more informal evaluation is not a retreat from rigor nor a simplification of what evaluators do. It requires increased attention, better observation, and more accurate and useful description.

In the evaluation of Title I programs, two masters must be served: the federal reporting mandate and local needs.

In Chicago this is becoming possible. Computerized processing of standardized test scores allows the reporting requirement to be met with a minimum of staff and time. The bulk of staff and time are used to gather information about program implementation and conduct at the local schools. This information is returned to school principals and program administrators several times yearly. In fiscal 1981 this feedback is to become almost instantaneous and is to occur continually. Program vendors are expected to become more active participants in the information flow. The Title I evaluation design facilitates this flow but also allows the retention of information for more traditional and formal end-of-year evaluation reports for schools, projects, and the overall program.

The Title I evaluation staff is working increasingly closely with principals and others--both at scheduled meetings and in informal consultations--to learn what information is desired and useful at the schools. Rules of procedure and conduct are being established jointly beforehand (although much progress remains to be made) and informal presentations of findings are being made at the close of each school visit and at frequent intervals to staff not located at the schools. It is expected that some principals will continue to resist these efforts, if only because their time is limited. Nevertheless, the emphasis on personal contact and direct involvement is beginning to produce improved instruction.

This emphasis on personal contact and involvement is the central tenet of the emerging rationale for the form of Title I

evaluation in Chicago. The evaluation is, in the informal, phenomenological sense, beginning to focus on two sets of persons. The first is composed of the student-teacher dyad as subject of the evaluation. It is their relationship which sets the limits to pupils' learning. The second set is composed of the school principal and the various administrators as the object of the evaluation. They have the power to promote or vitiate programs.

The more typical formal, empiricist approach slights both subject and object. Analysis of test scores and questionnaire responses replaces the student-teacher relationship. The search for "laws" and regularities of behavior overwhelms the interest of principals and administrators. It also ignores the changes over time in these interests and relationships. The reality in which these persons act fluctuates, sometimes dramatically, often in short spans of time. Nevertheless, that varying reality, much of which is political in essence, is the context in which decisions about Title I programs are made. Remaining aloof of this context can only bode ill for the success--in the sense of impact and relevance--of any evaluation.

NOTES

1. That the models ask no questions about the shape, content, or implementation of these programs is disquieting. Without knowledge of what a program looked like "on the ground," any evaluation must be of dubious value. The assignment of a result to a non-program is a real risk.

2. Although Chicago was not aware of the format of Title I evaluation in Cincinnati (Ahn, Barta and Rockwood 1976), there appear to be several similarities, particularly concerning the local school's perception of the role of the Title I evaluator. Title I evaluators need to develop greater levels of trust and rapport with local school staff (Felix 1978; Hamilton 1980) and school personnel must come to see the relevance of evaluation to their work (David 1978).

3. The reader sometimes gains the impression that analytic techniques are chosen more for their impressiveness than for their utility. The propensity in evaluation reports toward discussion of bias and better measurement is a consequence of this approach and sometimes seems to take precedence over the actual analysis.

REFERENCES

Ahn, U.R., Barta, M.B., & Rockwood, S.F. Localizing evaluation for public schools: Concept and practice. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

Bissell, J.S. Use of educational evaluation and policy studies by the California legislature. Educational Evaluation and Policy Analysis, 1979, 1(3), 29-3[illegible].

Bloom, B.S., Hastings, J.T., & Madaus, G.F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Burstein, L. & Miller, M.D. Alternative analytical models for identifying educational effects: Where are we? Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

David, J.L. Local uses of Title I evaluations. Menlo Park, CA: SRI International, July 1978.

Eisner, E.W. The educational imagination: On the design and evaluation of school programs. New York: MacMillan, 1979a.

Eisner, E.W. The use of qualitative forms of evaluation for improving educational practice. Educational Evaluation and Policy Analysis 1979b, 1(6),11-9.

Felix, J.L. Research and evaluation to improve instruction: The Cincinnati strategy. Educational Evaluation and Policy Analysis, 1979, 1(2), 57-62.

Florio, D.H., Behrmann, M.M., & Goltz, D.L. What do policy makers think of educational research and evaluation? Or do they? Educational Evaluation and Policy Analysis, 1979, 1 (6), 61-87.

Gamel, N.N., Tallmadge, G.K., Wood, C.T., & Brinkley, J.L. State ESEA Title I reports: Review and analysis of past reports, and development of a model reporting system and format. Mountain View, CA: RMC Research Corp., October 1975.

Getzels, J.W. & Guba, E.C. Social behavior and the administrative process. School Review, 1957, 65(4), 423-44.

Hamilton, G.E. Title I evaluation from the perspective of the local school. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Kearns, D.R. ESEA Title I evaluation model equivalence and comparability with real data. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

La Belle, T.J., Moll, L.C., & Weisner, T.S. Context-based educational evaluation: A participant research strategy. Educational Evaluation and Policy Analysis, 1979, 1(3), 85-94.

Linn, R.L. Validity of inferences based on the proposed Title I evaluation models. Educational Evaluation and Policy Analysis, 1979, 1(2), 23-32.

Linn, R.L. & Slinde, J.A. The determination of the significance of change between pre- and posttesting periods. Review of Educational Research, 1977, 47(1),121-50.

Schermerhorn, G.R. & Williams, R.G. An empirical comparison of responsive and preordinate approaches to program evaluation. Educational Evaluation and Policy Analysis, 1979, 1(3), 55-60.

Schwab, J.J. The practical: A language for curriculum. Washington: National Education Association, 1970.

Scriven, M. The methodology of evaluation. AERA Monograph Series on Curriculum Evaluation, 1967, 1, 39-83.

Stake, R.E. A theoretical statement of responsive evaluation. Studies in Educational Evaluation, 1976, 2(1), 19-22.

Tallmadge, G.K. & Wood, C.T. User's guide: ESEA Title I evaluation and reporting system. Mountain View, CA: RMC Research Corp., October 1976.

Wiley, D.E. Evaluation by aggregation: Social and methodological biases. Educational Evaluation and Policy Analysis, 1979, 1 (2), 41-5.

Wisler, C.E. & Anderson, J.K. Designing a Title I evaluation system to meet legislative requirements. Educational Evaluation and Policy Analysis, 1979, 1(2), 47-55.